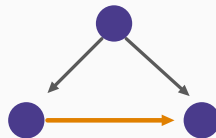


Inférence causale et machine learning

Chapitre 9 : Le contrôle synthétique



Une seule unité traitée

La construction

Les vérifications

Le lien avec le machine learning

Ce qu'il faut retenir

Une seule unité traitée

La situation

Une région adopte le programme, les autres non. Un pays réforme, ses voisins non.

Il n'y a **qu'une** unité traitée, observée sur de nombreuses périodes.

Pourquoi la double différence ne convient pas

Elle exige un groupe témoin dont la trajectoire aurait été **parallèle**.

Avec une seule unité traitée, quel témoin choisir ? Aucune région ne ressemble exactement à celle qui a réformé, et le choix d'un témoin unique est arbitraire — les résultats varient fortement selon la région retenue.

L'idée du contrôle synthétique

Plutôt que de choisir un témoin, on en **construit** un : une moyenne pondérée des unités non traitées, dont les poids sont choisis pour reproduire la trajectoire de l'unité traitée **avant** le traitement.

La construction

Comment les poids sont choisis

Le critère

On cherche les poids $w_j \geq 0$, de somme 1, qui minimisent l'écart entre l'unité traitée et la combinaison pondérée, **sur la période antérieure** au traitement.

On ajuste à la fois sur la variable de résultat et sur des caractéristiques prédictives.

Les deux contraintes

Les poids sont **positifs** et **somment à un**.

Elles interdisent l'extrapolation : le contrôle synthétique reste à l'intérieur de l'enveloppe des unités disponibles. C'est ce qui le distingue d'une régression, qui peut produire des combinaisons aberrantes.

Un contrôle synthétique typique

Région	Poids
A	0,42
B	0,31
C	0,18
D	0,09

Les vérifications

Comment juger un contrôle synthétique

Quatre contrôles

1. **Ajustement pré-traitement** — l'écart doit être faible sur toute la période antérieure.
2. **Placebos dans l'espace** — appliquer la méthode à chaque unité non traitée; l'effet estimé pour l'unité réellement traitée doit ressortir.
3. **Placebo dans le temps** — avancer artificiellement la date : rien ne doit apparaître.
4. **Retrait successif** — retirer chaque unité du réservoir et vérifier la stabilité.

Le test par placebos, et l'inférence

Avec une seule unité traitée, les tests statistiques usuels ne s'appliquent pas.

On construit donc une distribution de référence en traitant **chaque unité** du réservoir comme si elle avait été traitée. Si l'écart observé pour la vraie unité traitée est parmi les plus grands, on conclut à un effet.

C'est une **inférence par permutation**, adaptée à un cas où l'on ne dispose pratiquement d'aucun degré de liberté.

Les conditions de validité

Le lien avec le machine learning

La lecture prédictive

Le contrôle synthétique **prévoit** ce qu'aurait fait l'unité traitée sans traitement, à partir des autres unités.

C'est un problème de prévision, avec une contrainte d'interprétabilité : poids positifs, de somme un, et parcimonieux.

Ce que cela ouvre

On peut remplacer l'ajustement pondéré par n'importe quelle méthode d'*Apprentissage statistique* — régression pénalisée, forêt, boosting — pour construire le contrefactuel.

Ces variantes existent et prévoient souvent mieux. Elles perdent en revanche la lisibilité des poids, et donc une part de ce qui rendait la méthode convaincante.

C'est le premier point du cours où le machine learning entre réellement en jeu. Le chapitre 10 le fait de façon systématique.

Ce qu'il faut retenir

Cinq points

1. Le **contrôle synthétique** traite le cas d'une **seule** unité traitée.
2. Il **construit** un témoin par moyenne pondérée, ajustée sur la période antérieure.
3. Poids **positifs** et de **somme un** : pas d'extrapolation, et une solution parcimonieuse.
4. L'inférence se fait par **placebos** dans l'espace et dans le temps.
5. C'est un problème de **prévision du contrefactuel** — d'où la transition vers le machine learning.

La suite

Cinq stratégies d'identification ont été présentées. Le chapitre 10 montre ce que l'apprentissage automatique leur apporte — et ce qu'il ne leur apporte pas.